

- 1 -

Method for the Compression of Recordings of Ambient Noise,
Method for the Detection of Program Elements therein, and
Device therefor

5 Background of the Invention

10 The present invention refers to a method for the compression
of an electric audio signal which is produced in the process
of recording the ambient noise by means of an
electroacoustic transducer, more particularly a microphone.
Furthermore, the invention also refers to a device for
carrying out the method.

15 In the field of audience research, which also comprises the
acoustic perception of other media such as e.g. television,
recordings of the acoustic environment of a panelist in a
survey are used, i.e. the so-called hearing samples. The
storage of these hearing samples on portable magnetic tape
recorders is disclosed in US 5,023,929. The inconvenient of
20 this method is that the tape recorder is relatively large
although it is intended to be permanently carried by the
participant.

25 Consequently, it would be preferable to integrate the
hearing sample recorder or monitor in an appliance which is
normally worn or at least less visible. Such a possibility,
namely the integration into a wristwatch, is mentioned in
EP-A-0 598 682 to the applicant, this application being
hereby incorporated into the present specification.

30 However, the mentioned application does not indicate how the
hearing samples can be stored in the extremely narrow space
and with the very limited energy available in a wristwatch
or a similarly inconspicuous appliance over a considerable
35 period of time such as at least a week. Although the

- 2 -

specification mentions the need of compression procedures, known methods only are indicated.

Summary of the Invention

5

It is therefore an object of the present invention to provide a method for the compression of hearing samples which in particular allows to obtain a high compression with minimal efforts with the safe recognition of program elements being essentially conserved.

This object is attained by a method for the compression of an electric audio signal which is produced in the process of recording the ambient noise by means of an electroacoustic transducer, more particularly a microphone, wherein

- the amplitude of said audio signal or of a derived digital or analog signal is normalized to a first predetermined range D;
- said audio signal is mapped in the form of a non-linear mapping onto a second predetermined range of values W in order to obtain an emphasis of sensitive values; and
- the result is stored in an electronic memory in a digital form.

~~The further claims indicate preferred embodiments, devices~~
25 ~~for carrying out the method, and applications.~~

In the following, the same terminology as in EP-A-0 598 682 will be used. A hearing sample is basically a recording of the ambient noise e.g. by means of a microphone. In order to simplify the storage as well as the transmission to the evaluating center, however, it is preferred to have a succession of short recordings of the ambient noise or hearing samples which are recorded at certain times. Preferably, the recordings are effected at regular intervals of e.g. 1 minute, and have a constant duration of the order

- 3 -

of, for example, 4 seconds, the information of the time of the recordings being stored together with the hearing sample.

5 According to the invention, the hearing samples are finally stored in an electronic memory in a digitized form. According to the invention, in order to reduce the amount of data to be stored, a normalization of the hearing samples in their original form or in a derived form (filtered, limited
10 to selective frequency bands, digital or analog, etc.) to a predetermined range (of values or amplitudes) D and a subsequent nonlinear transformation on a second range W is effected whose result, which is limited to the range W, is then stored in an electronic memory. The range W may be
15 smaller or equal to D, but it is preferably substantially smaller.

Essentially, the non-linear transformation serves the purpose of amplifying sensitive areas of range D in such a
20 manner that the more significant information provided by a signal whose value is comprised in such a sub-range of D is emphasized in the result, i.e. its resolution is increased.

Preferred further developments of the invention are as
25 follows:

A: The nonlinear mapping is characterized by a decreasing slope dW/dD for increasing values in D, e.g. similar to the logarithmic function. Essentially, the range
30 of small values in D is thereby mapped onto a relatively larger range in W and thus emphasized, whereas relatively large values in D are mapped on a relatively small range in W only, i.e. their significance is attenuated.

- 4 -

5 B: The hearing samples are digitized immediately after recording (e.g. by a microphone) and analog processing (amplification; coarse filtering in preparation of the analog-digital conversion, etc.), resulting in a succession of numeric values. Each numeric value represents e.g. the momentary loudness of the ambient noise at a determined time.

10 Further processing is effected digitally by digital circuits, program controlled processors, or combinations thereof.

15 C: The amplitude or loudness values are transformed into energy values e.g. by squaring. The energy values are submitted to a low pass filtering and subsequently differentiated, the differentiation preferably being simulated by a difference calculus. The resulting energy variation values indicate the variation of the low-frequency proportion of the energy content in
20 time.

25 D: The group of the energy variation values of a hearing sample, or only a part thereof, is normalized with respect to the maximum value of the values within the (partial) group. For this purpose, the maximum value is determined and all values of the group are divided by this maximum value. Simultaneously, the normalized values are mapped on a given range of numbers corresponding to the range D, e.g. the numbers between
30 -128 and +127, so that the following arithmetic operations involve only integers. The number of values in these numerical ranges D is therefore preferably equal to powers of 2 (in the example: $256 = 2^8$ values) which are particularly advantageous in the
35 case of binary digital processing. In order to

- 5 -

perform this combination of normalizing and of
imaging, the values of a group are multiplied by a
factor which results from the division of the limit of
the numeric range (i.e. 128 in the example) by the
maximum value within the group.

E: The results of this step are again mapped on a
further, smaller range of values W, e.g. the numerical
range from 0 to 15 comprising $2^4 = 16$ numbers. On
account of the fixed and relatively small number of
values of the input data of this step, a so-called
look-up table may be used for this second mapping.

Overall, it follows from the preceding that each
numerical value of the hearing samples is reduced to a
relatively short binary number (of 4 bits in the
example).

F: Further optimizations are applied, such as e.g. taking
the mean value of a plurality of values, only the mean
value being further used. This also results in an
important reduction of the number of values to be
processed. On the digital level, such a filtering is
simulated by a convolution.

G: Before or after being digitized at the input, the
hearing sample is split into frequency bands or band
signals. In a known manner, digital filterings may be
effected by convolutions, and since the preferred
convolutions represent low pass filterings, it is
preferable to transmit less values to the following
processing stages than are used for the convolution,
preferably only one respective value.

Brief Description of the Drawings

The invention will be explained in more detail hereinafter by means of an exemplary embodiment and with reference to 5 figures.

Fig. 1 shows a block diagram of a monitor according to the invention;

10 Fig. 2 shows the division into frequency bands;

Fig. 3 shows the conversion into energy values and the differentiation;

15 Fig. 4 shows the "normalizing quantization".

Detailed Description of the Invention

Fig. 1 shows a block diagram of a monitor 1. It may e.g. be 20 intended to be integrated in a wristwatch, which is why monitor 1 comprises a clock circuit 2 which also serves as a time base for the signal processing, as well as a (liquid crystal) display 3. Commercially available components may be used for circuit 2 and display 3. A precise clock signal 25 is generated by a quartz 4 in conjunction with an oscillator circuit which is integrated in clock circuit 2. Since a highly precise timing is required for the synchronization of the hearing samples to the comparative samples, a temperature compensation is provided in addition. The 30 latter comprises a temperature sensor 5 which is connected to the clock circuit by means of an interface circuit 6. Interface circuit 6 essentially comprises an A/D converter.

Another important element for the monitor function is 35 wearing detector 7. It may essentially consist of a sensor

- 7 -

area on the wristwatch which detects the contact with the skin of the wearer. In the example, wearing sensor 7 is connected to clock circuit 2 by means of an interface circuit 8, which implies that the clock circuit is capable of providing the time indications with an additional mark from the wearing sensor. It is also conceivable to directly connect the wearing sensor to the proper monitor circuit, e.g. to digital signal processor 9.

10 The clock signals which are required for the signal processing, in particular for signal processor 9, are derived from the time base clock, which is taken from a connection 10 of quartz 4, by a PLL (phase locked loop) circuit 11. The time and the date as well as the mark from the wearing sensor, as the case may be, are transmitted from clock circuit 2 to digital signal processor 9 by a serial data connection 12.

The hearing samples are stored in a flash memory. It is an important advantage with respect to the present application that flash memories are capable of storing data in a non-volatile manner and of deleting them again without the need of particular measures. A bus 14 allowing to transmit both data and addresses serves to connect flash memory 13 and signal processor 9.

A multiplexer 16 is connected by a second serial connection. Depending on the operational condition, the multiplexer connects signal processor 9 to the recording unit of the hearing samples or to interface circuit 17 by means of which the data exchange with the evaluating center is effected.

The recording unit consists of a microphone 18 and a following A/D converter unit 19 which in addition to the proper A/D converter may comprise amplifiers, filters (anti-

5 Power supply 20 may be a battery (lithium cell) or the like. An accumulator in conjunction with a contactless charging system by means of electromagnetic induction or a photo cell is also conceivable.

25 As shown in the illustration, the hearing sample processing unit including signal processor 9 and the necessary accessory components (multiplexer 16, memory 13, clock generator consisting of PLL circuit 11 and quartz 10, etc.) may be composed of discrete components. In order to be
30 incorporated in a wristwatch, however, the functions must be integrated in as few components as possible, which may result in a single application specific circuit 30 in the extreme case. For example, signal processors of the TMS 320C5x series (manufacturer: Texas Instruments) may be used,
35 in which multiplexer 16 is already contained, inter alia,

- 9 -

and Flash RAMs of the type AM29LV800 (manufacturer: Amdahl) having a capacity of 8 MBit. Such a memory capacity and the application of the compression method for hearing sample data according to the invention as described hereinafter
5 allow to attain an uninterrupted operation of the monitor for approx. 7 days.

In view of energy consumption, it is advantageous if the hearing sample processing unit, more particularly signal
10 processor 9, is only periodically switched on. If e.g. one hearing sample per minute is taken, it is sufficient according to the processing method of the present invention to switch on the power supply of the signal processor for some seconds (less than 5, e.g. 4 seconds) only. For this
15 purpose, the power supply receives an on-signal 25 from clock circuit 2 during whose presence the hearing sample processing unit is supplied with current. A further reduction of the energy consumption is obtained by the fact that flash memory 13 is only supplied with the current
20 required for the storing process for a short time, 3 milliseconds at the end of each processed hearing sample recording being sufficient in the case of the above-suggested type. The signal 26 required therefor is generated by signal processor 9. The program controlling
25 the signal processor is contained in a separate program memory which may be integrated in the signal processor itself, so that the hearing sample processing operation can also be performed while flash memory 13 is off.

30 Hereinafter, the method for the processing of the hearing samples is described. After the recording of the ambient noise (microphone 18) and its analog-digital conversion according to known principles (A/D converter unit 19), a splitting into e.g. six frequency bands is performed (Fig.
35 2) which is effected by a hierarchical arrangement of low

0910239 062249

- 10 -

passes 30 - 35. The required high pass associated to each low pass is realized by a subtraction 36 - 41 of the output signals 42 - 47 from the respective input signals 48 - 53 of the low passes, the subtraction being effected by an addition of the inverted output signals 42 - 47 of low passes 30 - 35.

Low pass filters 30 to 35 are realized by a 19-digit convolution:

$$y_j = \sum_{i=0}^{18} a_i x_{j-i} \quad (1)$$

where

j : time index
 y_j : output value of the low pass filtering at the time j ;
 x_j : input value for low pass filtering at the time j ;
 a_i : coefficient of the convolution sequence;
 $a_0 \dots a_{18}$: [0.03, 0.0, -0.05, 0.0, 0.06, 0.0, -0.11, 0.0, 0.32, 0.50, 0.32, 0.0, -0.11, 0.0, 0.06, 0.0, -0.05, 0.0, 0.03]

In the course of the splitting into the frequency bands or band signals (54), a first data reduction is already effected in that only every second value out of each sequence of output values of the high and low pass filterings is transmitted to the following low resp. high pass stage or to outputs 54 by the switches 55. Overall, this already allows to obtain a reduction of the data volume to 1/8. With the division into six bands used in the example, this results in a slight overcompensation of the accompanying increase of the data volume by a factor six.

- 11 -

A criterion for the design of the filters is that one band may contain the contents of every other band in a clearly attenuated form at the most. A reduction to the half at least may be considered as clearly attenuated. Ideally, the
 5 bands only contain residual portions of directly adjacent bands, portions which are near or below the resolution of the digital numerical representation even. In the preferred digital realization, this aim is attained by low pass
 10 filtering (convolution) and subsequent subtraction of the filtered proportion from the input signal of the low pass filter.

The treatment of the band signals 54 resulting from the division into bands is identical in each band, Figs. 3 and 4
 15 showing the processing of only one band 56 in a representative manner.

Input signal 56, which is identical to output signal 54, is first squared in that it is supplied to the two inputs of a
 20 multiplier 57 in parallel. Except a proportionality factor, this squaring corresponds to a calculation of the energy content of the proportion of the ambient noise which is represented by signal 56. Energy values 58 are subjected to a low pass filtering. This filtering is realized by means
 25 of a convolution over 48 values:

$$y_j^e = \sum_{i=0}^{47} b_i x_{j-i}^e \quad (2)$$

where

30

j : time index of the y^e and x^e values;
 x_j^e : energy value 58 at the time j;

- 13 -

a hearing sample (the hearing sample duration used in this embodiment results in 32 energy difference values 64 per hearing sample in each frequency band).

5 In a reciprocal-computing and multiplication unit 71, the number 128 ($= 2^7$) is divided by the maximum value of the hearing sample and the result is supplied to an input 72 of a multiplier 73. The other input of multiplier 73 is then successively supplied with the energy difference values
10 64 among which the maximum value has been determined. For this purpose, the difference values 64 are temporarily stored in a FIFO buffer 75. The result of the multiplication in multiplier 73, whose values are comprised between -128 and +127, is converted by converter
15 76 into integers in the range D from 0 to 255, corresponding to a byte having 8 bits. These numbers are used as addresses in a look-up table (LUT) 77 where a number in the range $W = 0$ to 15, i.e. a four-digit binary number, is associated to each input value. The discrete mapping of 8-
20 bit numbers onto 4-bit numbers performed in LUT 77 is nonlinear and so designed that the resolution of small input numbers is finer than that of greater input values, i.e. that small input values are more emphasized. This may be referred to as a non-equidistant quantization.
25
The 4-bit values from output 78 are stored in flash memory 13 (Fig. 1).

The described normalized, non-equidistant quantization and
30 compression unit is provided for each band according to the illustration of Fig. 3, resulting in 4-bit values for a total of $32 \times 48 \times 8 = 12,288$ values per processing cycle which are recorded by the A/D converter at input 48 (Fig. 2). With an A/D conversion rate of 3,000 to 5,000
35 conversions per second, as provided by the currently

09102939.062398

- 14 -

available A/D converters of the lowest power consumption, this results in a hearing sample duration of approx. 2.5 to 4 s. With a supposed rate of one hearing sample per minute, the necessary memory capacity for the data amounts to 32×6
5 $\times 4 = 768$ bit/min or 1'105'920 bit/d. The indicated 8 Mbit memory thus allows to record approx. 7 days of uninterrupted operation of the monitor.

In view of a reduction of the required computing, all cited
10 calculations are effected by integer or fixed point arithmetic unless especially indicated, in particular an exponential representation of floating point numbers is avoided. The number of bits used for the representation of a number essentially depends on the used processor and on
15 the data length provided by the latter. The above-mentioned processor family TMS320C5x uses 16-bit arithmetic. The binary point for fixed point arithmetic is set in such a manner that the limited computing accuracy is optimally utilized in each processing step although the probability of
20 a data overflow is extremely low. Therefore, the binary point is set differently in the different processing steps. In the preferred embodiment of the band division, the least significant bit represents the value 2^{-16} for the filter coefficients and the value 2^0 for the data values. Energy
25 conversion and energy filtering are calculated by 32-bit integer arithmetic which is implemented as standard library function calls.

Prior to the storage in the flash memory or alternatively in
30 the evaluating center, usual compression methods may be additionally applied which allow restoration of the original data in an identical form when decompressed.

In preparation of the recognition of the program elements
35 which are possibly contained in the hearing samples, program

- 16 -

```

Apply same preliminary processing as to hearing samples;
FOR TimeShift := 1 TO MaxTimeShift STEP Timestep DO
  {Takes account of running inaccuracies of the timers by a step size of
  Timestep}
5   Calculate matching coefficient  $c_t$  with standard correlation for the
    actual time shift and assign result to the variable ActualMatch;
      IF (ActualMatch > OptimumMatch) DO
        OptimumMatch := ActualMatch;
        OptimumTimeShift := TimeShift;
10      OptimumStationaryUnit := Stationary Unit;
      ENDIF
    ENDFOR
  ENDFOR
15 IF (OptimumMatch > Threshold) DO
  RadioStation is recognized;
  The correct station is stored in the memory OptimumStationaryUnit
  ELSE
    None of the surveyed reference programs was heard at this time
20 ENDIF

```

In this procedure, only one of the radio programs registered in 'NumberOfStationaryUnits' is determined in the hearing sample of a monitor, namely the one which yields the highest probability (value of the variable 'OptimumMatch').

In particular, the optional, univocally reversible compression of the hearing samples processed according to the invention is reversed. This is followed by the initialization of 'OptimumMatch' to the lowest value which also indicates "no match", i.e. the wearer of the monitor has listened to none of the monitored programs.

The program samples of each stationary unit simultaneously recorded with the current hearing sample (loop "For

- 17 -

StationaryUnit:= 1 to NumberOfStationaryUnits ... EndDo" are loaded and processed in the same manner as the hearing sample. Due to subsisting small time shifts between the hearing samples and the program samples, the following

5 comparison is performed for a certain number 'MaxTimeShift' of assumed time shifts (loop "For TimeShift := 1 to MaxTimeShift ... Endfor"). The comparison is effected by a standard correlation of program and hearing sample data which are shifted forwards or backwards with respect to each

10 other according to the 'TimeShift' variable. In order to always allow a full correlation over all values of the hearing sample, the program samples are therefore recorded over a longer period per sample, the beginning being additionally set earlier in time by the corresponding

15 maximum time shift. Correspondingly, the length of the program sample is chosen in such a manner that the hearing sample is still completely contained in the program sample time even if the beginnings of the program sample and of the hearing sample are maximally displaced.

20

The normalized correlation is performed according to the following formula:

$$c_t = \frac{\sum_{i=1}^N (s_i m_{i-t})}{\sqrt{\sum_{i=1}^N (s_i)^2} \sqrt{\sum_{i=1}^N (m_{i-t})^2}} \quad (3)$$

25

where

t : time shift index (= 'TimeShift' in pseudocode);

N : number of correlated values, generally equal to

30 the number of values in a hearing sample;

i : time index;

s_i : hearing sample value at the time i;

- 18 -

m_{i-t} : program sample value at the time i , displaced by
 t time steps;
 c_t : correlation value for the time shift t : $-1 \leq c_t \leq$
 1 .

5

The c_t values for different t values and program samples are
 compared, and the greatest c_t value overall is stored along
 with the indications of the conditions in which it has been
 recorded. These indications consist of the time shift, the
 10 stationary unit, i.e. the program, and of the correlation
 value c_t itself.

If the so determined greatest c_t value is superior to a
 predetermined threshold value, the corresponding program is
 15 considered to be contained in the hearing sample. If the
 threshold value is not attained, it is assumed that no one
 of the programs was heard.

Since the correlation must be performed correspondingly
 20 often due to the considerable scope of time shifts (t resp.
 TimeShift), a simplified alternative is conceivable where
 the time intervals are treated with a coarser graduation.
 For those c_t values which exceed a predetermined threshold,
 the correlation is repeated with a more rugged method while
 25 taking account of all detected time shifts.

A suitable rugged correlation is

$$r_t = \frac{\sum_{i=1}^N |s_i - a * m_{i-t}|}{\sum_{i=1}^N |s_i|} \quad (4)$$

30

where

- 19 -

r_t : "rugged" correlation value;
 a : scaling factor which takes account of the
attenuation of the program signal with respect
to the hearing sample;

5 the remaining symbols corresponding to formula (3).

The procedure thus essentially uses absolute values both of
the deviation between the hearing sample and the scaled
program signal and of the hearing sample signal. The
10 scaling factor a is iteratively determined in such a manner
that the rugged correlation value r_t becomes minimal.
Compared to the normal correlation, large deviations are
less weighted in the rugged correlation, thus taking account
of statistical distributions of hearing sample values and of
15 program signal values and therefore resulting in better
recognition rates for real signals than the normal
correlation value c_t . In particular, individual hearing
samples with large deviations are less weighted.

20 Tests show that the described method not only eliminates or
at least strongly reduces known interference effects such as
secondary noise and time shifts but that damping (speakers,
transmission lines, general acoustic conditions) and echo as
well have only little influence on the recognition of a
25 program. It has been particularly surprising to find that
the program could often be detected in the hearing samples
even when the program element was inaudible. The
suppression of echo effects is attributed to the formation
of a temporal mean (filter 59), in particular, especially if
30 its time constant is chosen in such a manner as to be
greater than the echo times usually found in a normal
environment. A typically frequency-dependent (acoustic)
damping is compensated by the described suitable combination
of a division into frequency bands, a normalization to the
35 maximum value, and in taking into account of the damping by

09102939-062308

- 20 -

means of the scaling factor a in the calculation of r_t or by the calculation mode of c_t .

Modifications of the exemplary embodiment within the scope
5 of the invention are apparent to those skilled in the art.

According to the technological development, different components (signal processors, memories, etc.) may be used. Alternatives are conceivable in particular for the flash
10 memory, e.g. battery-backed up CMOS memories. The criteria, especially for portable monitors such as wristwatches, are an extended uninterrupted monitoring period and a minimal energy consumption. In certain circumstances it may be better to use a fast processing unit having a higher power
15 dissipation if the higher energy consumption with respect to a slower unit is more than compensated by only temporary operation with intermediate inactive pauses. Besides the complete shut-off, many components such as e.g. the TMS320C5xx also offer special power saving modes. Also, the
20 reduction of the clock rate of a fast unit often allows an important reduction of the energy consumption.

Depending on the used technology, different degrees of accuracy or numbers of digits of the binary numbers may be
25 used. In tests, a sufficiently safe program recognition has been obtained with 4-bit end results. It is also conceivable, however, to effect a reduction to 3 bits, or to provide a greater number, e.g. 6 bits, 7 bits, or 8 bits. Greater numbers of binary digits are possible in particular
30 if shorter wearing times are allowed or if memories of greater capacity become available.

In the case of higher numbers of digits of the end result, it may also be necessary to increase the number of digits in

- 21 -

the preceding steps to the number of digits of the end result at least.

Mostly, the exact values for the nonlinear mapping by table
5 77 as well as the threshold values for the weighting of the correlation values can only be determined empirically. Although a function similar to a logarithmization is preferred, other functions are possible. It is also conversely conceivable to emphasize the greater values in D
10 and to suppress the small values of the energy differences.

The factors and the number of digits of the convolutions may as well be chosen differently, and a different number of frequency bands into which the hearing samples are split is
15 possible. In particular, it is conceivable in the case of modified A/D conversion speeds, different settings with respect to echo and/or damping compensation, or modified hearing sample durations, to adapt low pass 59, e.g. by changing the number of tabs of the convolution.

It is also conceivable to perform the analog-digital conversion at a later stage of the compression, particularly if the corresponding analog circuits offer advantages with respect to the processing speed or the space consumption in
25 the monitor. In the extreme case, the digitization might be effected only immediately prior to the storage in the memory. If an analog signal is concerned, the term "digital value" in the description shall be replaced with e.g. the size or the amplitude of the signal.

With respect to the correlation, it is also possible to use only the part of the hearing samples which still lies within the corresponding program sample with the actual time shift t, e.g. if program and hearing samples of the same length
35 are recorded.

- 22 -

An alternative of the wearing sensor consists of using currently available motion sensors. A known embodiment contains a contact which switches between the open and the closed state on motion but remains in one of the two states in the absence of motion.

Glossary

- 10 Flash RAM RAM (see there) which also conserves data in case of power failure but allows faster storage and easier erasure than classic non-volatile memories (PROM/EPROM).
- RAM read/write memory
- 15 time index number of a digital value in the succession of values leaving the digitizer (A/D converter), mostly in relation to the beginning of a hearing sample, whose associated value has the time index 0.

20

- - - - -